On primary and secondary repetitions in words

Roman Kolpakov[a]

[a] *Moscow State University, Leninskie Gory, 119992 Moscow, Russia*

Abstract

Combinatorial properties of maximal repetitions (runs) in formal words are studied. We classify all maximal repetitions in a word as primary and secondary where the set of all primary repetitions determines all the other repetitons in the word. Essential combinatorial properties of primary repetitions are established.

1. Inroduction

Let $w = w[1]w[2]\dots w[n]$ be an arbitrary word. The length n of w and is denoted by $|w|$. A word $w[i]\cdots w[j]$, where $1 \le i \le j \le n$, is called a *factor* of w and is denoted by $w[i..j]$. Note that factors can be considered as fragments of the original word or as words themselves. So for factors we have two different notions of equality: factors can be equal as the same fragment of the original word or as the same word. To avoid this ambiguity, we will use two different notations: if two factors u and v are the same word (the same fragment of the original word) we will write $u = v$ ($u \equiv v$). For any $i = 1, \dots, n$ the factor $w[1..i]$ ($w[i..n]$) is called a *prefix* (a *suffix*) of w. A positive integer p is called a *period* of w if $w[i] = w[i+p]$ for each $i = 1, \dots, n-p$. We denote by $p(w)$ the minimal period of w and by $e(w)$ the ratio $|w|/p(w)$ which is called the *exponent* of w. A word is called *primitive* if its exponent is not an integer greater than 1.

By repetition in a word we mean any factor of exponent greater than or equal to 2. Repetitions are fundamental objects, due to their primary importance in word combinatorics [16] as well as in various applications, such as string matching algorithms [12, 2], molecular biology [14], or text compression [21]. The simplest and best known example of repetitions is factors of the form uu, where u is a nonempty word. Such repetitions are called *squares*. We will call the first (second) factor u of the square uu *the left (right) root* of this square. Avoiding ambiguity[1], by the *period* of a square we will mean the length of its roots. A square is called *primitive* if its roots are primitive. Primitive squares are a particular case of factors of the form $u^k = \underbrace{uu\dots u}_{k}$ where $k > 1$ and u is a nonempty primitive word. Such factor is called a *primitive integer power* with the root u. A primitive integer power is called *maximal* if it cannot be extended to the left or to the right in the word by at least one root. Note that any primitive integer power is contained in only one maximal integer power. In an analogous way, one can note that any repetition is contained in only one *maximal* repetition with the same minimal period which cannot be extended to the left or to the right in the word by at least one letter with preserving its minimal period. Maximal repetitions are usually called *runs* in the literature. Since runs contain all the other repetitions in a word, the set of all runs can be considered as a compact encoding of all repetitions in the word which has many useful applications (see, for example, [7]).

Questions concerning the maximum possible number of repetitions in words are actively investigated in the literature. In particular, it is shown in [1, 2] that the maximum possible number of primitive square and maximal integer powers in words of length n is $\Theta(n \log n)$. It is proved in [15] that, unlike the case of maximal integer powers, the maximum possible number $mrn(n)$ of runs in words of length n is $O(n)$ and,

Email address: foroman@mail.ru (Roman Kolpakov)

[1] Note that the period of a square is not necessarily the minimal period of this word.

moreover, the maximum possible sum $mex(n)$ of all runs in words of length n is also $O(n)$. Due to a series of papers [18, 19, 17, 3, 13, 4, 5] more precise upper bounds on $mrn(n)$ have been obtained. For the present time the best upper bound $1.029n$ on $mrn(n)$ is obtained in [5]. The problem of low bounds on $mrn(n)$ is considered in [9, 10, 20]. More precise bounds on $mex(n)$ have been also obtained in [3, 4, 8]. In particular, the best known bounds $mex(n) \leq 4.1n$ and $mex(n) > 2.035n$ are obtained in [8]. Analogical estimates for runs with exponent at least 3 are obtained in [6, 8].

Further we denote by $R(w)$ the set of all maximal repetitions in a word w. Let λ be a natural number. For maximal repetitions, in our opinion, one could make the two following natural conjectures:

1. The number of maximal repetitions with the minimal period not less than λ in the word w is upper bounded by $\varphi(\lambda)n$ where $\varphi(\lambda) \to 0$ as $\lambda \to \infty$.
2. The maximal number of maximal repetitions containing the same letter [2] in the word w is $o(n)$.

Unfortunately, both the conjectures are not true. As a counterexample, we can consider the word $w_k = (01)^k(10)^k$ of length $4k$. It is easy to check that $R(w_k) = \{r_1, r_2, \ldots, r_{k+2}\}$ where $r_1 = (01)^k$, $r_2 = (10)^k$, and $r_i = (1(01)^{k-3})^2$ for $i = 3, 4, \ldots, k+2$. Thus, for any $\lambda > 2$ the word w_λ contains $\lfloor \lambda/2 \rfloor = \Omega(|w_\lambda|)$ maximal repetitions with the minimal period not less than λ which contradicts conjecture 1. Moreover, the middle letters of w_k are contained in $k+1 = \Omega(|w_k|)$ different maximal repetitions from $R(w_k)$ which contradicts conjecture 2. However, one can easily observe that w_k has actually two "original" adjacent maximal repetitions r_1 and r_2 which "generate" all the other repetitions $r_3, r_4, \ldots, r_{k+2}$. This observation suggests that it would be possible to indicate in $R(w)$ a subset of repetitions which "generate" all the other maximal repetitions of w. In this paper we formally define the notion of generation of repetitions. In accordance with this notion, generated repetitions are called secondary and all the other maximal repetitions are called primary. Originally the notions of primary and secondary repetitions were introduced in [11] where they was used for space efficient search for maximal repetitions. In [11] some auxiliary combinatorial results for primary and secondary repetitions are also obtained. The notions of primary and secondary repetitions defined here are slightly different from the notions introduced in [11]. However, this difference is not crucial. Thus, in the present paper we continue the combinatorial investigations started in [11] for primary and secondary repetitions. In particular, we show that, unlike the case of all maximal repetitions, both conjectures 1 and 2 are true for primary repetitions. More precisely, we prove that in the word w the sum of exponents of all primary repetitions with the minimal period not less than λ and all secondary repetitions generated by these primary repetitions is $O(n/\lambda)$ which obviously implies that the number of primary repetitions with the minimal period not less than λ in the word w is also $O(n/\lambda)$. Moreover, we prove that the maximal number of primary repetitions which have the minimal period not less than λ and contain the same letter in the word w is $O(\log(n/\lambda))$ which obviously implies that conjecture 2 is also true for primary repetitions. Thus, the set of all primary repetitions which represent actually all repetitions in a word is more convenient for considering and treatment than the set of all maximal repetitions.

2. Auxiliary definitions and results

The results of the paper are based on the following well-known fact which is usually called *the periodicity lemma*.

Lemma 1. *If a word w has two periods p, q, and $|w| \geq p + q$, then $\gcd(p, q)$ is also a period of w.*

Using the periodicity lemma, it is easy to obtain

Proposition 2. *Let q be a period of a word w such that $|w| \geq 2q$. Then q is divisible by $p(w)$.*

We will use also the following evident fact.

[2] By the same letter we mean that letters in different positions of the word are different.

Proposition 3. *If two factors of a word have the same period q and are overlapped by at least q letters then q is a period of the union of these factors.*

Let $w = w[1]w[2]\dots w[n]$ be an arbitrary word. A repetition $r \equiv w[i..j]$ in w is called *maximal* if it satisfies the following conditions:

1. if $i > 1$, then $w[i-1] \neq w[i-1+p(r)]$,
2. if $j < n$, then $w[j+1-p(r)] \neq w[j+1]$.

In other words, a repetition in w with the minimal period p is maximal if its one letter extension in w (to the left or to the right) results in a factor with the minimal period $> p$. It is obvious that any repetition in a word is contained in only one maximal repetition with the same minimal period. We denote by $R(w)$ the set of all maximal repetitions in w. The following fact about maximal repetitions is a trivial consequence of Proposition 3.

Proposition 4. *The overlap of two non-separated different maximal repetitions with the same minimal period p is smaller than p.*

Proposition 4 obviously implies

Proposition 5. *Let $r' \equiv w[i'..j']$, $r'' \equiv w[i''..j'']$, $r''' \equiv w[i'''..j''']$ be different maximal repetitions in w with the same minimal period and $i' \leq i'' \leq i'''$. Then r' and r''' are not non-separated.*

3. Primary and secondary repetitions

Let r be a repetition in the word w. We call any factor of w which has the length $p(r)$ and is contained in r a *cyclic root* of r. Note that for any cyclic root u of r the word r is a factor of the word u^k where k is a large enough number. So it follows from the minimality of the period $p(r)$ that any cyclic root of r has to be a primitive word. Hence any two adjacent cyclic roots of r form a primitive square with the period $p(r)$ which is called a *cyclic square* of r. Two repetitions r' and r'' with the same minimal period p are called *cognate* if the words r' and r'' are factors of the same word u^k, where $|u| = p$ and k is a large enough number. It easy to see that cognate repetitions have the same set of distinct cyclic roots. For cognate repetitions we have the following statement which is proved in ([11], Lemma 1).

Lemma 6. *Let r', r'' be cognate repetitions with minimal period p in the word w. Then for any cyclic roots $u' \equiv w[i'_u..i'_u+p-1]$, $v' \equiv w[i'_v..i'_v+p-1]$ of r' and any cyclic roots $u'' \equiv w[i''_u..i''_u+p-1]$, $v'' \equiv w[i''_v..i''_v+p-1]$ of r'' such that $u' = u''$, $v' = v''$ an equality $i''_u - i'_u \equiv i''_v - i'_v \pmod{p}$ holds.*

Lemma 6 implies that there exists a residue class modulo p, such that, for any equal cyclic roots $u' \equiv w[i'_u..i'_u + p - 1]$ of r' and $u'' \equiv w[i''_u..i''_u + p - 1]$ of r'', the value $i'' - i'$ belongs to this class. We denote by $\sigma(r', r'')$ the minimal non-negative residue of this class. It is easy to see that cognate non-separated repetitions r', r'' are extended to the same maximal repetition if $\sigma(r', r'') = 0$. Hence

Proposition 7. *For any different cognate non-separated maximal repetitions r', r'' the value $\sigma(r', r'')$ is positive.*

We use also the following fact which is proved actually in ([11], Lemma 2) (here we present a shorter proof of this fact).

Lemma 8. *Let $r' \equiv w[i'..j']$, $r'' \equiv w[i''..j'']$ be cognate non-separated repetitions with minimal period p, and $v \equiv w[l..l+2q-1]$ be a primitive square with the period q such that $q \geq 2p$ and v is contained completely in $w[i'..j'']$. Than $i'' \leq l + q \leq j' + 1$.*

Proof. Denote respectively the roots $w[l..l+q-1]$ and $w[l+q..l+2q-1]$ of v by u' and u''. Suppose that $l+q < i''$. Note that in this case u' is contained completely in r', so p is a period of u'. Therefore, p is also a period of u''. If v is contained completely in r' than v has both periods p and q such that $|v| = 2q > p+q$. So by the periodicity lemma in this case v has also the period $\gcd(p,q)$ which contradicts the primitivity of roots of v. Thus we can suppose that $l+2q-1 > j'$. Let $j'+1-p \geq l+q$. Than both letters $w[j'+1-p]$ and $w[j'+1]$ are contained in u''. Since p is a period of u'', we obtain that $w[j'+1-p] = w[j'+1]$ which contradicts that r' is maximal. Now let $j'+1-p < l+q$. Taking into account that $j' \geq i''-1$, in this case we have

$$l+2q-1 > j'-p+q \geq j'+p \geq i''+p-1 > i''-1 \geq l+q,$$

so both letters $w[i''+p-1]$ and $w[i''-1]$ are contained in u''. Since p is a period of u'', we conclude that $w[i''+p-1] = w[i''-1]$ which contradicts that r'' is maximal. Thus $i'' \leq l+q$. The inequality $l+q \leq j'+1$ is proved by symmetrical way.

Let $r' \equiv w[i'..j']$, $r'' \equiv w[i''..j'']$ where $i' \leq i''$ be cognate non-separated repetitions from $R(w)$ with minimal period p. Then it follows from Proposition 4 that $i' < j'+1-p < i'' \leq j'+1 < j''$. We say that a repetition $r \equiv w[i..j]$ from $R(w)$ *is generated* by repetitions r' and r'' if the following conditions are valid:

1. $p(r) \geq 3p$;
2. $i' < i \leq j'$,
3. $i'' \leq j < j''$.

We will also say in this case that r' (r'') generates r from left (from right). If a repetition is generated by some repetitions from from $R(w)$ we call this repetition *secondary*. All repetitions from $R(w)$ which are not secondary are called *primary*. By $Rp(w)$ we denote the set of all primary repetitions in w, and by $Rs(w)$ we denote the set of all secondary repetitions in w.

Lemma 9. *Any secondary repetition is generated by only one pair of repetitions.*

Proof: Let a maximal repetition r be generated by a pair (r_1', r_1'') of repetitions with a minimal period p_1 and a pair (r_2', r_2'') of repetitions with a minimal period p_2 where $r_k' \equiv w[i_k'..j_k']$, $r_k'' \equiv w[i_k''..j_k'']$ for $k = 1, 2$. Consider in r an arbitrary cyclic square $v \equiv w[l..l+2p(r)-1]$. Since v is contained completely in $w[i_1'..j_1'']$ and $w[i_2'..j_2'']$, by Lemma 8, we have $i_k'' \leq l+p(r) \leq j_k'+1$ for $k = 1, 2$. Therefore, the left root $w[l..l+p(r)-1]$ of v is contained in both repetitions r_1' and r_2'. So r_1' and r_2' are overlapped by at least $p(r)$ letters where $p(r) > p_1 + p_2$. Moreover, the right root $w[l+p(r)..l+2p(r)-1]$ of v is contained in both repetitions r_1'' and r_2''. So r_1'' and r_2'' are also overlapped by at least $p(r)$ letters. Hence, by Proposition 4, we have $r_1' \equiv r_2'$ and $r_1'' \equiv r_2''$.

On the other hand, we can describe explicitly all repetitions generated by a given pair of repetitions.

Lemma 10. *Let a maximal repetition r in a word w be generated by a pair (r', r'') of repetitions with a minimal period p where $r' \equiv w[i'..j']$, $r'' \equiv w[i''..j'']$. Then $p(r) = \alpha p + \sigma(r', r'')$, and $r \equiv w[i''-p(r)..j'+p(r)]$, where α is an arbitrary integer satisfying the inequalities*

$$3 \leq \alpha < \frac{1}{p}\big(\min\{i''-i', j''-j'\} - \sigma(r', r'')\big). \tag{1}$$

Proof: Consider in r an arbitrary cyclic square $v \equiv w[l..l+2p(r)-1]$. Let $u' \equiv w[l..l+p-1]$ ($u'' \equiv w[l+p(r)..l+p(r)+p-1)$ be the prefix of length p in the left (right) root of v. By Lemma 8 we have $i'' \leq l+p(r) \leq j'+1$, so the left root of v is contained in r' and the right root of v is contained in r''. Hence u' is a cyclic root of r' and u'' is a cyclic root of r''. Since $u' = u''$, by Lemma 6 we obtain that $l+p(r)-l = p(r) \equiv \sigma(r', r'')$, i.e. $p(r) = \alpha p + \sigma(r', r'')$. Since $\sigma(r', r'') < p$, from $p(r) \geq 3p$ we have $\alpha \geq 3$. If α satisfies inequalities (1), then it is easy to note that all factors $w[l..l+2p(r)-1]$ such that $i'' \leq l+p(r) \leq j'+1$ are cyclic square of the same maximal repetition $w[i''-p(r)..j'+p(r)]$ which is generated by (r', r''). Thus $r \equiv w[i''-p(r)..j'+p(r)]$. Using Lemma 8, it is also not difficult to see that

if $\alpha \geq \frac{1}{p}\big(\min\{i''-i', j''-j'\} - \sigma(r', r'')\big)$, either $w[i'..j'']$ doesn't contain primitive squares with the period $\alpha p + \sigma(r', r'')$ or such squares are cyclic square of a repetition $w[i..j]$ which doesn't satisfy the conditions $i' < i$ or $j < j''$.

Corollary 11. *Any secondary repetition is generated by a pair of primary repetitions.*

Proof: Let $r = w[i..j]$ be a secondary repetition generated by a pair (r', r'') of repetitions with minimal period p in a word w where $r' \equiv w[i'..j']$, $r'' \equiv w[i''..j'']$. Then, by Lemma 10, we have

$$e(r) = \frac{2p(r) + \delta}{p(r)} = 2 + \frac{\delta}{p(r)} \tag{2}$$

where δ is the overlap of repetitions r', r''. Since $p(r) \geq 3p$ and $\delta < p$, due to Proposition 4, the equality (2) implies $e(r) < 7/3$. Thus the exponent of any secondary repetition is less than $7/3$. Consider now the repetition $r' = w[i'..j']$. Since $i = i'' - p(r)$ by Lemma 10, we have

$$|r'| \geq i'' - i' > i'' - i = p(r) \geq 3p.$$

Hence $e(r') > 7/3$. Similarly we can prove that $e(r'') > 7/3$. So neither r' nor r'' can be a secondary repetition.

Using Lemma 1 and Corollary refsecbyprim, we can easily compute all secondary repetitions from the set of all primary repetitions. So the set $Rp(w)$ represents actually all repetitions in w.

Corollary 12. *Any repetition r generates from left less than $e(r) - 2$ repetitions.*

Proof: It is easy to see from Proposition 5 that any maximal repetition r can have to the right only one maximal repetition r' non-separated and cognate with r. Thus all repetitions generated by r from left have to be generated by only one pair (r, r') of repetitions. From Lemma 10 we conclude that the number of repetitions generated by this pair is no more than the number of integer α such that $3 \leq \alpha < e(r)$ which is obviously less than $e(r) - 2$.

4. Main results

Further we consider pairs of integers (p, j) where $p > 0$. We will call such pairs *points*. For any two points (p', j'), (p'', j'') we say that the point (p', j') *covers* the point (p'', j'') if $p' \leq p'' \leq 4p'/3$ and $j' - (2p'/3) \leq j'' \leq j'$. By $V(p, j)$ we denote the set of all points covered by the point (p, j). Let $\mathcal{E}(w)$ be the set of all points (p, j) such that $1 \leq p \leq 2n/3$ and $1 \leq j \leq n$. For any repetition $r \equiv w[i..j]$ from $R(w)$ we denote by $\mathcal{P}(r)$ the set of all points $(p(r), i + kp - 1)$ of $\mathcal{E}(w)$ where k is an integer greater than or equal to 2 and $i + kp - 1 \leq j$. Note that $|\mathcal{P}(r)| = \lfloor e(r) - 1 \rfloor$, so for any repetition r the set $\mathcal{P}(r)$ is not empty. Moreover, from Proposition 4 we have

Proposition 13. *For any different repetitions r', r'' from $R(w)$ the sets $\mathcal{P}(r')$ and $\mathcal{P}(r'')$ are not intersected.*

We also use the following fact.

Proposition 14. *Two different points (p', j'), (p'', j'') of $\mathcal{E}(w)$ such that $p' = p''$ can not cover the same point.*

Proof. Let $p' = p''$. Then $j' \neq j''$. Assume without loss of generality that $j'' < j'$. Let the points (p', j'), (p'', j'') cover the same point (p, j). Then $j' - (2p'/3) \leq j \leq j'' < j'$. So the points (p', j'), (p'', j'') can not be contained in the same set $\mathcal{P}(r)$. On the other hand, if (p', j') and (p'', j'') are contained in the sets $\mathcal{P}(r')$, $\mathcal{P}(r'')$ for some different repetitions r' and r'' with the same minimal period $p' = p''$ then these repetitions have an overlap of length greater than or equal to $4p'/3$ which contradicts Proposition 4.

By $\mathcal{E}'(w)$ we denote the subset $\bigcup_{r \in Rp(w)} \mathcal{P}(r)$ of $\mathcal{E}(w)$. Note that, by Proposition 13, each point of $\mathcal{E}'(w)$ belongs to only one set $\mathcal{P}(r)$.

Our results are based on the following statement.

Lemma 15. *Three different points of $\mathcal{E}'(w)$ can not cover the same point.*

Proof. Let three different points (p_1, j_1), (p_2, j_2) and (p_3, j_3) of $\mathcal{E}'(w)$ cover the same point (p, j). Then, by Proposition 14, the numbers p_1, p_2 and p_3 have to be pairwise different. Assume without loss of generality that $p_3 < p_2 < p_1$. Note that in this case we have

$$p_3 < p_2 < p_1 \le p \le 4p_3/3.$$

For $k = 1, 2, 3$ let $r_k = w[s_k..t_k]$ be the primary repetition such that $(p_k, j_k) \in \mathcal{P}(r_k)$. Note that $p(r_k) = p_k$. Denote $j_k - 2p_k + 1$ by i_k. Note that the factor $w[i_k..j_k]$ is contained completely in r_k, so in r_k we can consider the conjugate cyclic roots $w[i_k..j_k - p_k]$ and $w[i_k + p_k..j_k]$. Denote respectively these roots by u'_k and u''_k. We also denote $p_2 - p_3$ by q. From $p_2 < 4p_3/3$ we have $q < p_3/3$. To prove the lemma, we consider separately the three following cases.

Case I. Let $j_2 \le j_3$. Note that in this case $i_2 < i_3$. First we prove that in this case $s_3 = i_3$, i.e. r_3 can not be extended with the same period to the left of $w[i_3]$. Assume that $s_3 \ne i_3$. Then, by definiton of $\mathcal{P}(r_3)$, the repetition r_3 has at least one cyclic root to the left of $w[i_3]$, i.e. the factor $w[i_3 - p_3..j_3]$ is contained completely in r_k. So p_3 is a period of $w[i_3 - p_3..j_3]$. Let $i_3 - p_3 \le i_2$. Then the factor $w[i_2..j_2]$ is contained in $w[i_3 - p_3..j_3]$. So $w[i_2..j_2]$ has both periods p_2 and p_3. Moreover, $|w[i_2..j_2]| = 2p_2 > p_2 + p_3$. Therefore, by the periodicity lemma $w[i_2..j_2]$ has the period $\gcd(p_2, p_3)$ which contradicts the primitivity of cyclic roots of r_2. Now let $i_3 - p_3 > i_2$. Then the overlap $w[i_3 - p_3..j_2]$ of factors $w[i_3 - p_3..j_3]$ and $w[i_2..j_2]$ has both periods p_2 and p_3. Since $j_3 - (2p_3/3) \le j \le j_2 \le j_3$ and $p_2 < 4p_3/3$, we have

$$|w[i_3 - p_3..j_2]| = 3p_3 - (j_3 - j_2) \ge 7p_3/3 > p_2 + p_3.$$

Therefore, by the periodicity lemma $w[i_3 - p_3..j_2]$ has the period $\gcd(p_2, p_3)$ which contradicts again the primitivity of cyclic roots of r_2. Thus, $s_3 = i_3$. Since $i_3 > 1$, it implies that $w[i_3 - 1] \ne w[i_3 + p_3 - 1]$. It is easy to see that $w[i_3 + p_3 - 1]$ is contained in u''_2. So from $u'_2 = u''_2$ we have $w[i_3 + p_3 - 1] = w[i_3 + p_3 - 1 - p_2] = w[i_3 - q - 1]$. Thus $w[i_3 - 1] \ne w[i_3 - q - 1]$. Denote by v the overlap $w[i_3 + p_3..j_2]$ of u''_2 and u''_3. Taking into account $j_2 \ge j \ge j_3 - (2p_3/3)$, we obtain

$$|v| = j_2 - (j_3 - p_3) \ge \frac{p_3}{3} > q.$$

Moreover, since $u'_2 = u''_2$ and $u'_3 = u''_3$, we have

$$v = w[i_3..j_2 - p_3] = w[i_3 + p_3 - p_2..j_2 - p_2] \tag{3}$$

which implies that q is a period of v. For case I we consider separately subcases $i_1 < i_3 - q$ and $i_1 \ge i_3 - q$.

Subcase I.1. Let $i_1 < i_3 - q$. Since $j_3 - (2p_3/3) \le j \le j_1$ and $p_1 \le 4p_3/3$, the relation $j_1 - p_1 \ge j_3 - 2p_3 = i_3 - 1$ is valis. Thus in this case we have $i_1 \le i_3 - q - 1 < 1_3 - 1 \le j_1 - p_1$. So both symbols $w[i_3 - q - 1]$ and $w[i_3 - 1]$ are contained in u'_1. Since $u'_1 = u''_1$, we obtain $w[i_3 - q - 1] = w[i_3 + p_1 - q - 1]$ and $w[i_3 - 1] = w[i_3 + p_1 - 1]$. Therefore, $w[i_3 + p_1 - q - 1] \ne w[i_3 + p_1 - 1]$. Using the inequalities $p_1 > p_2$ $p_1 \le 4p_3/3$, we obtain $i_3 + p_1 - q - 1 \ge i_3 + p_2 - q = i_3 + p_3$ and $i_3 + p_1 - 1 = j_3 - 2p_3 + p_1 \le j_3 - (2p_3/3) \le j \le j_2$. Thus both symbols $w[i_3 + p_1 - q - 1]$ and $w[i_3 + p_1 - 1]$ are contained in v. So $w[i_3 + p_1 - q - 1] \ne w[i_3 + p_1 - 1]$ contradicts the fact that q is a period of v. So this subcase is impossible.

Subcase I.2. Let $i_1 \ge i_3 - q$. Note that in this case $j_3 < j_1$. Consider the factor $v' \equiv w[i_3 - q..j_2 - p_3]$. Note from (3) that q is a period of v'. Moreover, $|v'| = |v| + q > 2q$. So v' is a repetition. Let q' be the minimal period of v'. Note that q' is a divisor of q by Proposition 2. Let $\hat{v}' \equiv w[i'..j']$ be the maximal repetition containing v'. Then we consider separately subcases $j' < j_3 - p_3$ and $j' \ge j_3 - p_3$.

Subcase I.2.a. Let $j' < j_3 - p_3$. Since the repetition $\hat{v}'$ is maximal, we have $w[j' + 1] \ne w[j' + 1 - q']$. It follows from the inequalities $j' \ge j_2 - p_3$, $j_2 \ge j \ge j_3 - (2p_3/3)$ and $q' \le q < p_3/3$ that $j' + 1 - q' > i_3$. Thus $i_3 \le j' + 1 - q' < j' + 1 \le j_3 - p_3$, i.e. both symbols $w[j' + 1 - q']$ and $w[j' + 1]$ are contained in u'_3. Since $u'_3 = u''_3$, we obtain that the symbols $w[j' + 1 + p_3]$ and $w[j' + 1 + p_3 - q']$ contained in u''_3 are also different. It follows from inequalities $j' \ge j_2 - p_3$, $j_2 \ge j \ge j_1 - (2p_1/3)$ and $q' \le q < p_3/3 < p_1/3$ that

$j'+1+p_3-q' > j_1-p_1$. On the other hand, since $w[j'+1+p_3]$ is contained in u''_3, we have $j'+1+p_3 \le j_3 < j_1$. Thus both symbols $w[j'+1+p_3]$ and $w[j'+1+p_3-q']$ are contained in u''_1. Since $u'_1 = u''_1$, we conclude that the symbols $w[j'+1+p_3-p_1]$ and $w[j'+1+p_3-q'-p_1]$ contained in u'_1 are also different. The inequality $p_3 < p_1$ implies that $j'+1+p_3-p_1 \le j'$. On the other hand, since $w[j'+1+p_3-q'-p_1]$ is contained in u'_1, we obtain

$$j'+1+p_3-q'-p_1 \ge i_1 \ge 1_3 - q \ge i'.$$

Thus both symbols $w[j'+1+p_3-p_1]$ and $w[j'+1+p_3-q'-p_1]$ are contained in $\hat{v}'$ which contradicts the fact that q' is a period of $\hat{v}'$. So this subcase is also impossible.

Subcase I.2.b. Let $j' \ge j_3 - p_3$. In this case u'_3 is contained in $\hat{v}'$, so q' is a period of u'_3. Since $|u'_3| = p_3 > 3q \ge 3q'$, using Proposition 2, it is easy to see that q' has to be the minimal period of u'_3. Thus, u'_3 is a repetition with the minimal period q'. So u''_3 is also a repetition with the minimal period q'. Let $\hat{v}'' \equiv w[i''..j'']$ be the maximal repetition containing u''_3. If $\hat{v}' \equiv \hat{v}''$, then $u'_3u''_3$ is contained in $\hat{v}'$, so q' is the minimal period of $u'_3u''_3$. Applying Proposition 2 to $u'_3u''_3$, we obtain that in this case q' is a divisor of p_3, so q' is a period of r_3 which contradicts the fact that p_3 is the minimal period of r_3. Thus $\hat{v}' \not\equiv \hat{v}''$. It is obvious that the repetitions $\hat{v}'$, $\hat{v}''$ are non-separated and cognate. Taking into account $i' \le i_3 - q$, $s_3 = i_3$, and $j' \ge j_3 - p_3$, we also have $i' < s_3 \le j'$. Note that, obviously, $i_1 < j_3 - p_3$. Consider the factor $v'_1 = w[i_1..j_3-p_3]$. The inequlities $i_3 - q \le i_1$ and $p_1 > p_2$ imply that $j_1 - p_1 > j_3 - p_3$, so v'_1 is contained in u'_1. Therefore, $u'_1 = u''_1$ implies $v'_1 = v''_1$ where $v''_1 \equiv w[i_1+p_1..j_3+p_1-p_3]$. It follows from $i' \le i_3 - q \le i_1$ and $j' \ge j_3 - p_3$ that v'_1 is also contained in $\hat{v}'$, so q' is a period of v'_1. Hence q' is also a period of v''_1. From $j_1 - p_1 > j_3 - p_3$ and $p_1 > p_3$ we have $i_1 + p_1 > i_3 + p_3$ and $j_3 + p_1 - p_3 > j_3$, so the overlap of v''_1 and u''_3 is $w[i_1+p_1..j_3]$. The inequlities $j_3 \ge j \ge j_1 - (2p_1/3)$ imply that the length of this overlap is no less than $p_1/3 > q \ge q'$. Hence, using Proposition 3, we obtain that q' is the minimal period of $w[i_3+p_3..j_3+p_1-p_3]$. So $w[i_3+p_3..j_3+p_1-p_3]$ is contained in $\hat{v}''$. Thus

$$j'' \ge j_3 + p_1 - p_3 > j_3 + p_2 - p_3 = j_3 + q.$$

Therefore, if $t_3 \ge j_3 + q$, then both numbers q' and p_3 are periods of the factor $w[i_3+p_3..j_3+q]$ and, moreover, the length of this factor is $p_3 + q$, i.e. is no less than $p_3 + q'$. Hence, by the periodicity lemma, in this case $w[i_3+p_3..j_3+q]$ has the period $\gcd(q', p_3)$ which contradicts the primitivity of cyclic roots of r_3. Thus, $t_3 < j_3 + q$, i.e. $t_3 < j''$. On the other hand, we have, obviously, $t_3 \ge i_3 + p_3 \ge i''$. Recall also that $p_3 > 3q \ge 3q'$. Summing up the inequalities proved above, we obtain that r_3 is generated by the repetitions $\hat{v}'$ and $\hat{v}''$, i.e. r_3 is a secondary repetition which contradicts $r_3 \in Rp(w)$. Thus, Case I is impossible.

Case II. Let $j_3 < j_2$ and $i_3 > i_2$. In this case we consider separately the three following subcases: $j_3 - p_3 > j_2 - p_2$, $j_3 - p_3 = j_2 - p_2$, and $j_3 - p_3 < j_2 - p_2$.

Subcase II.1. Let $j_3 - p_3 > j_2 - p_2$. Denote for convenience the root u''_3 by v. Note that in this subcase v is contained completely in u''_2. Thus, from $u'_2 = u''_2$ and $u'_3 = u''_3$ we obtain

$$v = w[i_3+p_3-p_2..j_3-p_2] = w[i_3..j_3-p_3]. \tag{4}$$

So q is a period of v. Moreover, $|v| = p_3 > 3q$. Thus, v is a repetition, and by Proposition 2 the minimal period q' of this repetition is a divisor of q. Denote by v' the factor $w[i_3+p_3-p_2..j_3-p_3]$. From (4) we have that v' is also a repetition with the minimal period q'. Let $s_3 \ne i_3$, i.e. r_3 has at least one cyclic root to the left of $w[i_3]$. Then v' is contained in r_3, so v' has both periods q' and p_3, and, moreover, $|v'| = p_3 + q \ge p_3 + q'$. Therefore, by the periodicity lemma v' has the period $\gcd(p_3, q')$ which contradicts the primitivity of cyclic roots of r_3. Thus, $s_3 = i_3$. Since $i_3 > 1$, it implies $w[i_3-1] \ne w[j_3-p_3]$. Since $j_3 - p_3 > j_2 - p_2$, the letter $w[j_3-p_3]$ is contained in u''_2, so $w[j_3-p_3] = w[j_3-p_3-p_2]$. Hence $w[i_3-1] \ne w[j_3-p_3-p_2] = w[i_3-q-1]$. In this subcase we consider separately the two following subcases.

Subcase II.1.a. Let $i_1 \le j_3 - p_3 - p_2$. From inequalities $j_3 - (2p_3/3) \le j \le j_1$ and $p_1 \le 4p_3/3$ we have that $j_1 - p_1 \ge j_3 - 2p_3 = i_3 - 1$. Thus

$$i_1 \le j_3 - p_3 - p_2 < i_3 - 1 \le j_1 - p_1,$$

i.e. both symbols $w[i_3-1]$ and $w[j_3-p_3-p_2]$ are contained in u'_1. So $w[i_3-1]=w[i_3+p_1-1]$ and $w[j_3-p_3-p_2]=w[j_3+p_1-p_3-p_2]$. Thus $w[i_3+p_1-1]\neq w[j_3+p_1-p_3-p_2]$. Using $p_1>p_2$, we obtain that $j_3+p_1-p_3-p_2\geq j_3+1-p_3=i_3+p_3$. On the other hand, the inequality $p_1\leq 4p_3/3$ implies $i_3+p_1-1<i_3+2p_3-1=j_3$. Thus we have that

$$i_3+p_3\leq j_3+p_1-p_3-p_2=i_3+p_1-q-1<i_3+p_1-1<j_3$$

i.e. both letters $w[i_3+p_1-1]$ and $w[j_3+p_1-p_3-p_2]$ are contained in v. Therefore, since $j_3+p_1-p_3-p_2=i_3+p_1-q-1$, the relation $w[i_3+p_1-1]\neq w[j_3+p_1-p_3-p_2]$ contradicts the fact that q is a period of v. So this subcase is impossible.

Subcase II.1.b. Let $i_1>j_3-p_3-p_2$. Consider the maximal repetitions $\hat{v}'\equiv w[i'..j']$ and $\hat{v}''\equiv w[i''..j'']$ containing respectively v' and v'' with the minimal period q'. By the same way as in subcase I.2.b we can prove that $\hat{v}'\not\equiv\hat{v}''$. Moreover, it is obvious that the repetitions $\hat{v}'$, $\hat{v}''$ are non-separated and cognate. Denote by v'_1 the factor $w[i_1..j_3-p_3]$. Note that v'_1 is contained in v', so q' is a period of v'_1. It follows from $i_1>j_3-p_3-p_2$ and $p_1>p_2$ that $j_1-p_1>j_3-p_3$. So v'_1 is contained in u'_1. Therefore, $u'_1=u''_1$ implies $v'_1=v''_1$ where $v''_1\equiv w[i_1+p_1..j_3+p_1-p_3]$. So q' is also a period of v''_1. Since $j_1-(2p_1/3)\leq j\leq j_3$, we have $j_1-p_1\leq j_3-(p_1/3)$. On the other hand, from $p_1>p_3$ we have $j_3+p_1-p_3>j_3$. Thus, the length of the overlap $w[i_1+p_1..j_3]$ of v and v''_1 is not less than $p_1/3$, i.e. is greater than q'. Hence, by Proposition 3, we obtain that q' is the minimal period of $w[i_3+p_3..j_3+p_1-p_3]$. So $w[i_3+p_3..j_3+p_1-p_3]$ is contained in $\hat{v}''$. Then, by the same way as in subcase I.2.b we can show that $i''\leq t_3<j''$. We have also that

$$i'\leq i_3+p_3-p_2=i_3-q<i_3=s_3\leq j_3-p_3\leq j'.$$

Thus, as in subcase I.2.b, we obtain that r_3 is generated by $\hat{v}'$ and $\hat{v}''$, i.e. r_3 is a secondary repetition which contradicts $r_3\in Rp(w)$. So subcase II.1 is impossible.

Subcase II.2. Let $j_3-p_3=j_2-p_2$. Then from $u'_2=u''_2$ and $u'_3=u''_3$ we obtain that q is a period of u'_2 and u''_2. Moreover, taking into account $|u'_2|=|u''_2|=p_2>p_3>3q$ and Proposition 2, we have that u'_2 and u''_2 are repetitions, and the minimal period q' of these repetitions is a divisor of q. Consider again the maximal repetitions $\hat{v}'\equiv w[i'..j']$ and $\hat{v}''\equiv w[i''..j'']$ with the minimal period q' containing respectively u'_2 and u''_2. By the same way as in subcase I.2.b we can prove that $\hat{v}'\not\equiv\hat{v}''$. If $s_3\neq i_3$ then u'_2 is contained completely in r_3, so u'_2 has both periods q and p_3. Since $|u'_2|=p_2=p_3+q$, using the periodicity lemma, we obtain in this case that u'_2 has the period $\gcd(p_3,q)$, so u'_3 has also the period $\gcd(p_3,q)$ which contradicts the primitivity of cyclic roots of r_3. Thus, $s_3=i_3$. Therefore,

$$i'\leq i_2<i_3=s_3\leq j_3-p_3=j_2-p_2\leq j'.$$

If $t_3\geq j_2$ then u''_2 is contained completely in r_3, so in this case we can also obtain a contradiction to the primitivity of cyclic roots of r_3. Hence

$$i''\leq i_2+p_2=i_3+p_3\leq j_3\leq t_3<j_2\leq j''.$$

It is also obvious that $\hat{v}'$, $\hat{v}''$ are non-separated and cognate. Thus, taking into account the inequalities proved above, we obtain in this subcase that r_3 is generated by $\hat{v}'$ and $\hat{v}''$ which contradicts $r_3\in Rp(w)$.

Subcase II.3. Let $j_3-p_3<j_2-p_2$. Denote by v' and v'' the factors $w[i_2..j_3-p_3]$ and $w[i_3+p_3..j_3+q]$ respectively. From $u'_2=u''_2$ and $u'_3=u''_3$ we have

$$w[i_2+q..j_3-p_3]=w[i_2+p_2..j_3]=w[i_2..j_3-p_2],$$

so q is a period of v'. Since u'_3 is contained in u'_2, by the same way we have

$$u''_3\equiv w[i_3+p_3..j_3]=u'_3=w[i_3+p_2..j_3+q],$$

so q is also a period of v''. Since $|v'|,|v''|>p_3>3q$, we obtain that v', v'' are cognate repetitions, and the minimal period q' of these repetitions is a divisor of q.

First we prove that $s_3 = i_3$. Assume that $s_3 \neq i_3$, i.e $s_3 \leq i_3 - p_3$. If $s_2 \neq i_2$, i.e. $s_2 \leq i_2 - p_2 < i_3 - p_3$, then the factor $w[i_3 - p_3..j_3]$ of length $3p_3$ has both periods p_3 and p_2. Since $p_2 + p_3 < 7p_3/3 < 3p_3$, by the periodicity lemma we obtain in this case that this factor has the period $\gcd(p_3, p_2)$ which contradicts the primitivity of cyclic roots of r_2. Thus, $s_2 = i_2$. Let $\hat{v}' \equiv w[i'..j']$ be the maximal repetition containing the repetition v', and $\hat{v}''$ be the maximal repetition containing the repetition v''. By the same way as in subcase I.2.b we can prove that $\hat{v}' \not\equiv \hat{v}''$. We consider separately the three following subcases.

Subcase II.3.a. Let $i_1 < i_2$. Then $i_2 > 1$, so $s_2 = i_2$ implies that $w[i_2 - 1] \neq w[j_2 - p_2]$. It follows from $j_3 - p_3 < j_2 - p_2$ that the letter $w[j_2 - p_2]$ is contained in u''_3, so $w[j_2 - p_2] = w[j_2 - p_2 - p_3] = w[i_2 + q - 1]$. Thus, $w[i_2 - 1] \neq w[i_2 + q - 1]$. Note that

$$p_1 \leq \frac{4}{3}p_3 < p_3 + \frac{1}{3}p_2 = \frac{4}{3}p_2 - q.$$

Using this estimation together with $j_1 \geq j \geq j_2 - (2p_2/3)$, we obtain $j_1 - p_1 > j_2 + q - 2p_2 = i_2 + q - 1$. Thus, $i_1 \leq i_2 - 1 < i_2 + q - 1 < j_1 - p_1$, i.e. both letters $w[i_2 - 1]$, $w[i_2 + q - 1]$ are contained in u'_1. Therefore, $w[i_2 - 1] = w[i_2 + p_1 - 1]$ and $w[i_2 + q - 1] = w[i_2 + p_1 + q - 1]$. Hence $w[i_2 + p_1 - 1] \neq w[i_2 + p_1 + q - 1]$. Using $p_1 > p_2$, we have $i_2 + p_1 - 1 \geq i_2 + p_2$, so $i_2 + p_1 - 1 > i_3 + p_3$. On the other hand, using $p_1 \leq 4p_3/3 < 2p_3$, we have

$$i_2 + p_1 + q - 1 < i_2 + 2p_3 + q - 1 < i_3 + 2p_3 + q - 1 = j_3 + q.$$

Thus, both letters $w[i_2+p_1-1]$, $w[i_2+p_1+q-1]$ are contained in v''. Therefore, $w[i_2+p_1-1] \neq w[i_2+p_1+q-1]$ contradicts the fact that q is a period of v''. So this subcase is impossible.

Subcase II.3.b. Let $i_1 \geq i_2$ and $j' < j_2 - p_2$. Since $\hat{v}'$ is maximal, we have $w[j' + 1] \neq w[j' + 1 - q']$. From $j_3 - p_3 \leq j' < j_2 - p_2$ and $q' \leq q < p_3$ we have also that $i_3 < j' + 1 - q' < j' + 1 \leq j_2 - p_2$, i.e. both letters $w[j'+1]$, $w[j'+1-q']$ are contained in u'_2. So $w[j'+1] = w[j'+p_2+1]$ and $w[j'+1-q'] = w[j'+p_2+1-q']$. Thus $w[j'+p_2+1] \neq w[j'+p_2+1-q']$. Note that in this subcase $j_1 > j_2$ and $j'+p_2 < j_2$, so $j'+p_2+1 < j_1$. On the other hand, we have $j' \geq j_3-p_3$, so $j'+p_2+1 > j_3+p_2-p_3 = j_3+q \geq j_3+q'$. Hence $j'+p_2+1-q' > j_3$. Taking into account $j_3 \geq j \geq j_1 - (2p_1/3) > j_1 - p_1$, we obtain

$$j_1 - p_1 < j' + p_2 + 1 - q' < j' + p_2 + 1 < j_1.$$

Thus, both letters $w[j' + p_2 + 1 - q']$ and $w[j' + p_2 + 1]$ are contained in u''_1. Hence $w[j' + p_2 + 1 - q'] = w[j' + p_2 + 1 - q' - p_1]$ and $w[j' + p_2 + 1] = w[j' + p_2 + 1 - p_1]$. So $w[j' + p_2 + 1 - q' - p_1] \neq w[j' + p_2 + 1 - p_1]$. Since $p_2 < p_1$, we have $j' + p_2 + 1 - p_1 \leq j'$. On the other hand, since $w[j' + p_2 + 1 - q' - p_1]$ is contained in u'_1, we have also $j' + p_2 + 1 - q' - p_1 \geq i_1 \geq i_2 \geq i'$. Thus, both letters $w[j' + p_2 + 1 - q' - p_1]$ and $w[j' + p_2 + 1 - p_1]$ are contained in $\hat{v}'$. So $w[j' + p_2 + 1 - q' - p_1] \neq w[j' + p_2 + 1 - p_1]$ contradicts the fact that q' is a period of $\hat{v}'$. Therefore, this subcase is also impossible.

Subcase II.3.c. Let $j' \geq j_2 - p_2$. Then u'_2 is contained completely in $\hat{v}'$, so q' is a period of u'_2. It follows from $j_3 - p_3 < j_2 - p_2$ and $p_2 < 4p_3/3 < 2p_3$ that $i_3 - p_3 < i_2$, so u'_2 is contained completely in r_3, i.e. p_3 is also a period of u'_2. Moreover, $|u'_2| = p_2 = p_3 + q \geq p_3 + q'$. Therefore, by the periodicity lemma this factor has the period $\gcd(p_3, q')$ which contradicts the primitivity of the root u'_3 contained in u'_2.

Since all the considered subcases are impossible, we conclude that $s_3 = i_3$. Then, analogously to subcase II.2, one can prove that r_3 is generated by $\hat{v}'$ and $\hat{v}''$ which contradicts $r_3 \in Rp(w)$. Thus, Case I is also impossible.

Case III. Let $i_3 \leq i_2$. In this case we consider separately the subcases $s_2 = i_2$ and $s_2 \neq i_2$.

Subcase III.1. Let $s_2 = i_2$. Denote by v the overlap $w[i_2..j_3 - p_3]$ of u'_2 and u'_3. It follows from $j_3 \geq j \geq j_2 - (2p_2/3)$ and $q < p_2/3$ that $|v| \geq (p_2/3) + q > 2q$. Since $u'_2 = u''_2$ and $u'_3 = u''_3$, we have

$$v = w[i_2 + p_2..j_3 + q] = w[i_2 + p_3..j_3], \tag{5}$$

so q is a period of v. Thus, v is a repetition, and by Proposition 2 the minimal period q' of this repetition is a divisor of q. Therefore, using again (5), we obtain that $w[i_2 + p_3..j_3 + q]$ is also a repetition with the minimal period q'. We denote this repetition by v'' and consider separately the subcases $i_1 < i_2$ and $i_1 \geq i_2$.

Subcase III.1.a. Let $i_1 < i_2$. Note that in this subcase $i_2 > 1$, so $s_2 = i_2$ implies that $w[i_2-1] \neq w[j_2-p_2]$. It follows from $i_3 \leq i_2$, $p_3 < p_2$ and $j_3 \geq j \geq j_2 - (2p_2/3)$ that $i_3 + p_3 \leq j_2 - p_2 < j_3$, i.e. the letter $w[j_2-p_2]$ is contained in u_3''. So $w[j_2-p_2] = w[j_2-p_2-p_3] = w[i_2+q-1]$. Thus $w[i_2-1] \neq w[i_2+q-1]$. Note that

$$\frac{4}{3}p_2 - q = \frac{1}{3}p_2 + p_3 > \frac{4}{3}p_3 \geq p_1.$$

Therefore, $j_1 \geq j \geq j_2 - (2p_2/3)$ implies that $j_1 - p_1 > j_2 + q - 2p_2 = i_2 + q - 1$. On the other hand, we have $i_1 \leq i_2 - 1$. Thus, both letters $w[i_2-1]$ and $w[i_2+q-1]$ are contained in u_1'. So $w[i_2-1] = w[i_2+p_1-1]$ and $w[i_2+q-1] = w[i_2+p_1+q-1]$. Therefore, $w[i_2+p_1-1] \neq w[i_2+p_1+q-1]$. From $p_1 > p_2$ we have $i_2+p_1-1 \geq i_2+p_2 > i_2+p_3$. On the other hand, using $p_1 < 4p_2/3$ and $j_3 \geq j \geq j_2 - (2p_2/3)$, we obtain

$$i_2 + p_1 + q - 1 = j_2 + p_1 + q - 2p_2 < j_2 + q - \frac{2}{3}p_2 \leq j_3 + q.$$

Thus, both letters $w[i_2+p_1-1]$ and $w[i_2+p_1+q-1]$ are contained in v''. Therefore, since q' is a divisor of q, the inequality $w[i_2+p_1-1] \neq w[i_2+p_1+q-1]$ contradicts the fact that q' is a period of v''. So this subcase is impossible.

Subcase III.1.b. Let $i_1 \geq i_2$. Denote by $\hat{v}' \equiv w[i'..j']$ the maximal repetition contaning the repetition v and by $\hat{v}'' \equiv w[i''..j'']$ the maximal repetition contaning the repetition v''. Let $\hat{v}' \equiv \hat{v}''$. Then $j' \geq j_3 + q$, so u_2' is contained completely in $\hat{v}'$. Therefore, q' is the minimal period of u_2', so q' is also the minimal period of u_2''. Since u_2'' is overlapped with $\hat{v}'$ by at least $|v|$ letters where $|v| > 2q > q'$, by Proposition 3 we obtain in this case that q' is the minimal period of $u_2'u_2''$. So, by Proposition 2, q' is a divisor of p_2 which contradicts the primitivity of cyclic roots of r_2. Thus, $\hat{v}' \not\equiv \hat{v}''$. By Proposition 4, $\hat{v}'$ and $\hat{v}''$ can not be overlapped by greater than or equal to q' letters, so $i'' \leq i_2 + p_3$ and $q' \leq q$ imply that $j' < j_2 - p_2$. Since the repetition $\hat{v}'$ is maximal, we have $w[j'+1-q'] \neq w[j'+1]$. It follows from $j'+1 \geq i_2 + |v| > i_2 + 2q \geq i_2 + 2q'$ and $j' < j_2 - p_2$ that $i_2 < j'+1-q' < j'+1 \leq j_2 - p_2$, i.e. both letters $w[j'+1-q']$ and $w[j'+1]$ are contained in u_2'. Therefore, $w[j'+1-q'] = w[j'+p_2+1-q']$ and $w[j'+1] = w[j'+p_2+1]$. Thus, $w[j'+p_2+1-q'] \neq w[j'+p_2+1]$. From $i_1 \geq i_2$ we obtain $j_1 > j_2$. Therefore, since $w[j'+p_2+1]$ is contained in u_2'', we have $j'+p_2+1 < j_1$. On the other hand, $j' \geq j_3 - p_3$ implies that $j'+p_2+1 > j_3+q$, so

$$j' + p_2 + 1 - q' > j_3 + q - q' \geq j_3 \geq j \geq j_1 - \frac{2p_1}{3} > j_1 - p_1.$$

Thus, both letters $w[j'+p_2+1-q']$ and $w[j'+p_2+1]$ are contained in u_1''. Therefore, $w[j'+p_2+1-q'] = w[j'+p_2+1-p_1-q']$ and $w[j'+p_2+1] = w[j'+p_2+1-p_1]$. Thus, $w[j'+p_2+1-p_1-q'] \neq w[j'+p_2+1-p_1]$. It follows from $p_1 > p_2$ that $j'+p_2+1-p_1 \leq j'$. On the other hand, since $w[j'+p_2+1-p_1-q']$ is contained in u_1', we have $j'+p_2+1-p_1-q' \geq i_1$, so $j'+p_2+1-p_1-q' \geq i_2 \geq i'$. Thus, both letters $w[j'+p_2+1-p_1-q']$ and $w[j'+p_2+1-p_1]$ are contained in $\hat{v}'$. Hence $w[j'+p_2+1-p_1-q'] \neq w[j'+p_2+1-p_1]$ contradicts the fact that q' is a period of $\hat{v}'$. Thus, subcase III.1 is impossible.

Subcase III.2. Let $s_2 \neq i_2$, i.e. $s_2 \leq i_2 - p_2$. Then, analogously to case I, one can prove that $s_3 = i_3$. Denote respectively by u' and u'' the factors $w[i_3-q..j_3-p_3]$ and $w[i_3+p_3..j_3+q]$. It is easy to see that $j_3 \geq j \geq j_2 - (2p_2/3) > j_2 - p_2$ implies $i_3 - q > i_2 - p_2$, i.e. $i_3 - q > s_2$, and $i_3 \leq i_2$ implies $j_3 + q < j_2$, i.e. $j_3 + q < t_2$. Thus, u' and u'' are contained completely in r_2, so u' and u'' are cyclic roots of r_2. Note that if one considers respectively u' and u'' instead of u_2' and u_2'', this subcase is identical to subcase II.2. Hence, by the same way as in subcase II.2, we can prove that in this subcase the repetition r_3 is secondary. This contradiction to $r_3 \in Rp(w)$ completes the proof of Lemma 15.

Further we assign to each point (p,j) the weight $\rho(p,j) = 1/p^2$, and for any finite set A of points we define

$$\rho(A) = \sum_{(p,j)\in A} \rho(p,j) = \sum_{(p,j)\in A} \frac{1}{p^2}.$$

Let λ be a positive integer. By $\mathcal{E}_\lambda(w)$ ($\mathcal{E}'_\lambda(w)$) we denote the set of all points (p,j) from $\mathcal{E}(w)$ ($\mathcal{E}'(w)$) such that $p \geq \lambda$. Using Lemma 15, we prove the following

Corollary 16. $|\mathcal{E}'_\lambda(w)| = O\left(\frac{n}{\lambda}\right)$.

Proof. It is obvious that for any point (p,j) from $\mathcal{E}'_\lambda(w)$ the set $V(p,j)$ is contained in $\mathcal{E}_\lambda(w)$. On the other hand, by Lemma 15, each point of $\mathcal{E}_\lambda(w)$ can not be covered by more than two points of $\mathcal{E}'_\lambda(w)$. Therefore,

$$\sum_{(p,j)\in\mathcal{E}'_\lambda(w)} \rho(V(p,j)) \leq 2\rho(\mathcal{E}_\lambda(w)) = 2\left(\sum_{(p,j)\in\mathcal{E}_\lambda(w)} \frac{1}{p^2}\right) = 2n\left(\sum_{\lambda\leq p\leq 2n/3} \frac{1}{p^2}\right).$$

Using the evident inequality $\frac{1}{p^2} < \int_{p-\frac{1}{2}}^{p+\frac{1}{2}} \frac{1}{x^2}\,dx$, we estimate

$$\sum_{\lambda\leq p\leq 2n/3} \frac{1}{p^2} < \sum_{p=\lambda}^{\infty} \frac{1}{p^2} < \int_{\lambda-\frac{1}{2}}^{\infty} \frac{1}{x^2}\,dx = \frac{1}{\lambda-\frac{1}{2}}.$$

Thus,

$$\sum_{(p,j)\in\mathcal{E}'_\lambda(w)} \rho(V(p,j)) < \frac{2n}{\lambda-\frac{1}{2}}.$$

On the other hand, for any point (p,j) we can also estimate $\rho(V(p,j))$:

$$\begin{aligned}\rho(V(p,j)) &> \frac{2p}{3}\sum_{p\leq i\leq 4p/3} \frac{1}{i^2} > \frac{2p}{3}\int_p^{\lfloor 4p/3\rfloor+1} \frac{1}{x^2}\,dx \\ &> \frac{2p}{3}\int_p^{4p/3} \frac{1}{x^2}\,dx = \frac{2p}{3}\cdot\frac{1}{4p} = \frac{1}{6}.\end{aligned}$$

So $\sum_{(p,j)\in\mathcal{E}'_\lambda(w)} \rho(V(p,j)) > \frac{1}{6}|\mathcal{E}'_\lambda(w)|$. Therefore,

$$|\mathcal{E}'_\lambda(w)| < 6\left(\sum_{(p,j)\in\mathcal{E}'_\lambda(w)} \rho(V(p,j))\right) < \frac{12n}{\lambda-\frac{1}{2}} = O\left(\frac{n}{\lambda}\right).$$

Let $Rp_\lambda(w)$ be the set of all repetitions from $Rp(w)$ with the minimal period greater than or equal to λ. It is obvious that $\mathcal{E}'_\lambda(w) = \bigcup_{r\in Rp_\lambda(w)} \mathcal{P}(r)$. Therefore, since all the sets $\mathcal{P}(r)$ for $r \in Rp_\lambda(w)$ are non-empty and pairwise disjoint by Proposition 13, the bound $|Rp_\lambda(w)| \leq ||\mathcal{E}'_\lambda(w)|$ takes place. Thus, Corollary 16 implies

Theorem 17. $|Rp_\lambda(w)| = O\left(\frac{n}{\lambda}\right)$.

Corollary 16 allows actually to strengthen this result. Let $Rs_\lambda(w)$ be the set of all secondary repetitions generated by repetitions from $Rp_\lambda(w)$. Denote by $exp_\lambda(w)$ the sum $\sum_{r\in Rp_\lambda(w)} e(r)$ of the exponents of all repetitions from $Rp_\lambda(w)$ and by $exs_\lambda(w)$ the sum $\sum_{r\in Rs_\lambda(w)} e(r)$ of the exponents of all repetitions from $Rs_\lambda(w)$. Then we have

Theorem 18. $exp_\lambda(w) + exs_\lambda(w) = O\left(\frac{n}{\lambda}\right)$.

Proof. By Lemma 9 each repetition r from $Rs_\lambda(w)$ can be corresponded to the repetition from $Rs_\lambda(w)$ which generates r from left. Hence $|Rs_\lambda(w)| \leq \sum_{r\in Rp_\lambda(w)} e(r) - 2$, due to Corollary 12. From the proof of Corollary 11 we can also conclude that the exponent of any secondary repetition is less than 3. Therefore,

$$exs_\lambda(w) < 3\cdot|Rs_\lambda(w)| \leq \sum_{r\in Rp_\lambda(w)} 3e(r) - 6.$$

Thus,

$$exp_\lambda(w)+exs_\lambda(w)<\sum_{r\in Rp_\lambda(w)}4e(r)-6=4\cdot\left(\sum_{r\in Rp_\lambda(w)}e(r)-\frac{3}{2}\right).$$

Using $|\mathcal{P}(r)|=\lfloor e(r)-1\rfloor$, we can estimate $e(r)-\frac{3}{2}<\frac{3}{2}|\mathcal{P}(r)|$, so $\sum_{r\in Rp_\lambda(w)}e(r)-\frac{3}{2}<\frac{3}{2}|\mathcal{E}'_\lambda(w)|$. Therefore, $exp_\lambda(w)+exs_\lambda(w)<6\cdot|\mathcal{E}'_\lambda(w)|$. Hence Theorem 18 follows immediately from Corollary 16.

Let $clp_\lambda(w,i)$ where $i=1,2,\ldots,n$ be the number of repetitions from $Rp_\lambda(w)$ which contain the letter $w[i]$, and $clp_\lambda(w)=\max_i clp_\lambda(w,i)$. The value $clp_\lambda(w)$ can be also estimated by Lemma 15.

Theorem 19. $clp_\lambda(w)=O(\log\frac{n}{\lambda})$.

Proof. Consider the number i such that $clp_\lambda(w)=clp_\lambda(w,i)$. Denote by R' the set of all repetitions from $Rp_\lambda(w)$ which contain $w[i]$. We correspond each repetition r from R' to some point $(p(r),j_r)$ of $\mathcal{P}(r)$ in the following way. If in $\mathcal{P}(r)$ there exists at least one point $(p(r),j)$ such that $j\geq i$ then j_r is the minimal number j such that $(p(r),j)\in\mathcal{P}(r)$ and $j\geq i$. Othervise j_r is the maximal number j such that $(p(r),j)\in\mathcal{P}(r)$. It is easy to note that in this case $i-p(r)<j_r<i+2p(r)$. Therefore, for any repetition r from R' we have that $V(p(r),j_r)$ is contained completely in the set of all points (p,j) such that

$$\lambda\leq p\leq\frac{2n}{3},\qquad i-\frac{5p}{3}<j<i+2p.$$

Denote this set by Ω. By Proposition 13 different repetitions from R' correspond to different points, and by Lemma 15 each point of W can not be covered by more than two different points corresponding to repetitions from R'. Thus,

$$\sum_{r\in R'}\rho(V(p(r),j_r))\leq 2\rho(\Omega).$$

Using the evident inequality $\frac{1}{p}<\int_{p-\frac{1}{2}}^{p+\frac{1}{2}}\frac{1}{x}\,dx$, we estimate $\rho(\Omega)$:

$$\begin{aligned}\rho(\Omega)&=\sum_{p=\lambda}^{\lfloor 2n/3\rfloor}\sum_{i-(5p/3)<j<i+2p}\frac{1}{p^2}<\sum_{p=\lambda}^{\lfloor 2n/3\rfloor}\frac{11p}{3}\cdot\frac{1}{p^2}=\frac{11}{3}\sum_{p=\lambda}^{\lfloor 2n/3\rfloor}\frac{1}{p}\\&<\frac{11}{3}\int_{\lambda-\frac{1}{2}}^{\lfloor 2n/3\rfloor+\frac{1}{2}}\frac{1}{x}\,dx<\frac{11}{3}\int_{\lambda-\frac{1}{2}}^{n}\frac{1}{x}\,dx=\frac{11}{3}\ln\frac{n}{\lambda-\frac{1}{2}}.\end{aligned}$$

On the other hand, it is shown in the proof of Corollary 16 that $\rho(V(p,j))>1/6$ for any point (p,j). So $\sum_{r\in R'}\rho(V(p(r),j_r))>|R'|/6$. Thus,

$$|R'|<6\sum_{r\in R'}\rho(V(p(r),j_r))\leq 12\rho(\Omega)<44\ln\frac{n}{\lambda-\frac{1}{2}}=O(\log\frac{n}{\lambda}).$$

Since $|R'|=clp_\lambda(w)$, we conclude $clp_\lambda(w)=O(\log\frac{n}{\lambda})$.

Thus, unlike the case of all maximal repetitions, only a logarithmic number of primary repetitions in a word can contain the same letter.

5. Conclusion

In the paper we define secondary repetitions as generated repetitions r satisfying the condition $p(r)\geq 3p$ where p is the minimal period of the generating repetitions. At the same time we suppose that the factor 3 in this condition is "conventional", i.e. we conjecture that for any natural $k\geq 3$ after replacing this condition by $p(r)\geq kp$ Theorems 17, 18 and 19 will remain true.

In the introduction we give an example of word which has many secondary repetitions. However, the total number of runs in this word is relatively small in comparison with the maximum possible number of runs in a word. This observations allows to make the conjecture that the words with the maximum possible number of runs have no secondary repetitions, i.e. $mrn(n)$ coincides with the maximum possible number of primary repetitions in words of length n.

Acknowledgments

This work is partially supported by Russian Foundation for Fundamental Research (Grant 11-01-00508).

References

[1] M. Crochemore, An optimal algorithm for computing the repetitions in a word, *Inform. Process. Lett.* **12** (1981), 244–250.

[2] M. Crochemore, W. Rytter, Squares, cubes, and time-space efficient string searching, *Algorithmica* **13** (1995), 405–425.

[3] M. Crochemore, L. Ilie, Analysis of maximal repetitions in strings, *Lecture Notes in Comput. Sci.* **4708** (2007), 465–476.

[4] M. Crochemore and L. Ilie, Maximal repetitions in strings, *J. Comput. Syst. Sci.* **74(5)** (2008), 796–807.

[5] M. Crochemore, L. Ilie, and L. Tinta, Towards a solution to the "runs" conjecture, *Lecture Notes in Comput. Sci.* **5029** (2008), 290–302.

[6] M. Crochemore, C. Iliopoulos, M. Kubica, J. Radoszewski, W. Rytter, T. Walen, On the maximal number of cubic runs in a string, *Lecture Notes in Comput. Sci.* **6031** (2010), 227–238.

[7] M. Crochemore, C. Iliopoulos, M. Kubica, J. Radoszewski, W. Rytter, T. Walen, Extracting powers and periods in a string from its runs structure, *Lecture Notes in Comput. Sci.* **6393** (2010), 258–269.

[8] M. Crochemore, M. Kubica, J. Radoszewski, W. Rytter, T. Walen, On the maximal sum of exponents of runs in a string, *Lecture Notes in Comput. Sci.* **6460** (2011), 10–19.

[9] F. Franek, R. Simpson, W. Smyth, The maximum number of runs in a string, *Proceedings of 14th Australasian Workshop on Combinatorial Algorithms* (2003), 26–35.

[10] F. Franek, Q. Yang, An asymptotic lower bound for the maximal number of runs in a string, *Int. J. Found. Comput. Sci.* **19(1)** (2008), 195–203.

[11] L. Gasieniec, R. Kolpakov, I. Potapov, Space efficient search for maximal repetitions, *Theoret. Comput. Sci.* **339** (2005), 35–48.

[12] Z. Galil, J. Seiferas, Time-space optimal string matching, *J. Comput. System Sci.* **26(3)** (1983), 280–294.

[13] M. Giraud, Not so many runs in strings, *Lecture Notes in Comput. Sci.* **5196** (2008), 232–239.

[14] D. Gusfield, Algorithms on Strings, Trees, and Sequences, Cambridge University Press, 1997.

[15] R. Kolpakov, G. Kucherov, On Maximal Repetitions in Words, *J. Discrete Algorithms* **1(1)** (2000), 159–186.

[16] M. Lothaire, Combinatorics on Words, volume 17 of *Encyclopedia of Mathematics and Its Applications*, Addison Wesley, 1983.

[17] S. Puglisi, J. Simpson, W. Smyth, How many runs can a string contain?, *Theoret. Comput. Sci.* **401** (2008), 165–171.

[18] W. Rytter, The number of runs in a string: improved analysis of the linear upper bound, *Lecture Notes in Comput. Sci.* **3884** (2006), 184–195.

[19] W. Rytter, The Number of Runs in a String, *Inform. and Comput.* **205(9)** (2007), 1459–1469.

[20] J. Simpson, Modified Padovan words and the maximum number of runs in a word, *Australasian J. of Comb.* **46** (2010), 129–145.

[21] J. Storer, Data compression: methods and theory, Computer Science Press, Rockville, MD, 1988.